EE5907/EE5027 : Probability Review: Solutions

Exercise 2.6

(a) According to Bayes Rule,

$$\vec{P}(H|e_1, e_2) = \frac{P(H, e_1, e_2)}{P(e_1, e_2)} = \frac{P(e_1, e_2|H)P(H)}{P(e_1, e_2)} \tag{1}$$

thus (ii) is sufficient for calculation

(b) Given $E_1 \perp E_2|H$, $P(e_1, e_2|H) = P(e_1|H)P(e_2|H)$
From (a), we have

$$\vec{P}(H|e_1, e_2) = \frac{P(e_1, e_2|H)P(H)}{P(e_1, e_2)} \tag{2}$$

From $E_1 \perp E_2|H$, we have

$$\vec{P}(H|e_1, e_2) = \frac{P(e_1|H)P(e_2|H)P(H)}{P(e_1, e_2)} \tag{3}$$

Eq. (3) corresponds to terms in (i). In addition, we can calculate $P(e_1, e_2)$ by $\sum_H (P(e_1, e_2|H)P(H))$, so (iii) is also sufficient.
To conclude, (i),(ii),(iii) are all sufficient.

Exercise 2.7

Proof by counter example:

(I) Let X_1 and X_2 be outcomes of independent coin toss (1 means head, 0 means tails). $X_3 = X_1 \oplus X_2$, where $\oplus$ is XOR operator. $p(X_3|X_1, X_2) \neq p(X_3)$ since X_1 and X_2 determines X_3 deterministically, so X_1, X_2, X_3 are not mutually independent. However, $p(X_3|X_1) = p(X_3)$, $p(X_3|X_2) = p(X_3)$, so X_1, X_2, X_3 are pairwise independent.

Exercise 2.8

Proof

($\Rightarrow$) Given $X \perp Y|Z$, we have $p(x,y|z) = p(x|z)p(y|z)$. Let $g(x,z) = p(x|z)$ and $h(y,z) = p(y|z)$, then $p(x,y|z) = g(x,z)h(y,z)$.

($\Leftarrow$) Suppose $p(x,y|z) = g(x,z)h(y,z)$. Integrate both sides over x (or summation if x is discrete)

$$\begin{aligned}\int p(x,y|z)dx &= \int g(x,z)dx \times h(y,z) \\ \implies p(y|z) &= G(z)h(y,z), \quad (4)\end{aligned}$$

where $G(z) = \int g(x,z)dx$.

Integrate both sides over y (or summation if y is discrete)

$$\begin{aligned}\int p(x,y|z)dy &= g(x,z) \times \int h(y,z)dy \\ \implies p(x|z) &= g(x,z)H(z), \quad (5)\end{aligned}$$

where $H(z) = \int h(y,z)dy$

Finally, let's integrate with respect to both x and y:

$$1 = \int\int p(x,y|z)dxdy = \int\int g(x,z)h(y,z)dxdy \quad (6)$$

$$= \int g(x,z)dx \int h(y,z)dy = G(z)H(z) \quad (7)$$

Therefore

$$\begin{aligned}p(x,y|z) = g(x,z)h(y,z) &= \frac{p(x|z)}{G(z)}\frac{p(y|z)}{H(z)} \text{ using Eq. (4) and Eq. (5)} \\ &= p(x|z)p(y|z) \text{ using Eq. (7)} \quad (8)\end{aligned}$$

EE5907/EE5027 Week 2: MLE + MAP: Solutions

Exercise 3.1

The likelihood is given by

$$p(\mathcal{D}|\theta) = \theta^{N_1}(1-\theta)^{N_0} \tag{1}$$

Hence the log-likelihood is given by

$$\log p(\mathcal{D}|\theta) = N_1 \log\theta + N_0 \log(1-\theta) \tag{2}$$

To optimize the log-likelihood, we get

$$\underset{\theta}{\operatorname{argmax}}\, p(\mathcal{D}|\theta) = \underset{\theta}{\operatorname{argmax}}(N_1 \log\theta + N_0 \log(1-\theta)) \tag{3}$$

Differentiating with respect to θ and set to 0, we get:

$$\begin{aligned}
&\frac{N_1}{\theta} - \frac{N_0}{1-\theta} = 0 \\
&\implies N_1(1-\theta) = N_0\theta \\
&\implies \theta = \frac{N_1}{N_1 + N_0} \\
&\implies \theta = \frac{N_1}{N}
\end{aligned}$$

Hence, $\hat{\theta}_{MLE} = \frac{N_1}{N}$

Exercise 3.6

The Poisson distribution can be represented as:

$$\mathcal{D} = (x_1, x_2, \cdots, x_n)\,, \mathcal{D} \sim Poi\,(\lambda) \tag{4}$$

The likelihood is given by

$$p(\mathcal{D}|\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \tag{5}$$

To optimize the log-likelihood, we get

$$
\begin{aligned}
\hat{\lambda}_{MLE} &\triangleq \underset{\lambda}{\operatorname{argmax}} \log \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\
&= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\
&= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \left(-\lambda + x_i \log \lambda - \log x_i! \right) \\
&= \underset{\lambda}{\operatorname{argmax}} \left(-n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \sum_{i=1}^{n} \log x_i! \right) \\
&= \underset{\lambda}{\operatorname{argmax}} \left(-n\lambda + \sum_{i=1}^{n} x_i \log \lambda \right)
\end{aligned}
$$

Differentiating with respect to λ and set to 0, we get:

$$
\begin{aligned}
&- n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0 \\
\implies &\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i
\end{aligned}
$$

Exercise 3.7

a. Multiply the likelihood by the conjugate prior given in the question, we get the following posterior:

$$
\begin{aligned}
&p(\lambda|\mathcal{D}) \propto p(\mathcal{D}|\lambda)p(\lambda) \propto e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \lambda^{a-1} e^{-\lambda b} \\
\implies &p(\lambda|\mathcal{D}) \propto \frac{1}{\prod_{i=1}^{n} x_i!} e^{-(n+b)\lambda} \lambda^{a-1+\sum_{i=1}^{n} x_i} \\
\implies &p(\lambda|\mathcal{D}) \propto \lambda^{a-1+\sum_{i=1}^{n} x_i} e^{-(n+b)\lambda} \\
\implies &p(\lambda|\mathcal{D}) = Ga\left(\lambda |\, a + \sum_{i=1}^{n} x_i, n+b \right)
\end{aligned}
$$

b. Given the mean of Gamma distribution $Ga\left(a, b\right)$ is $\frac{a}{b}$, we can get the mean of $p(\lambda|\mathcal{D})$ to be

$$
\bar{\theta} = \frac{a + \sum_{i=1}^{n} x_i}{n+b} \tag{6}
$$

Given that $a \to 0$ and $b \to 0$, we have

$$\lim_{a\to 0, b\to 0} \frac{a + \sum_{i=1}^{n} x_i}{n+b} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Hence, the posterior mean converges to the ML solution.

Exercise 3.12

The posterior of the Bernoulli

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

if $\theta = 0.5$,

$$\begin{aligned} & p(\mathcal{D}|\theta)p(\theta) = 0.5^{N+1} \\ \implies & \log p(\mathcal{D}|\theta)p(\theta) = (N+1)\log 0.5 \end{aligned}$$

if $\theta = 0.4$,

$$\begin{aligned} & p(\mathcal{D}|\theta)p(\theta) = 0.4^{N_1} 0.6^{N-N_1} 0.5 \\ \implies & \log p(\mathcal{D}|\theta)p(\theta) = N_1 \log 0.4 + (N - N_1)\log 0.6 + \log 0.5 \end{aligned}$$

if $\theta =$others,

$$p(\mathcal{D}|\theta)p(\theta) = 0$$

For 0.5 to win out over 0.4,

$$\begin{aligned} & (N+1)\log 0.5 > N_1 \log 0.4 + (N - N_1)\log 0.6 + \log 0.5 \\ \implies & N \log \frac{0.5}{0.6} > N_1 \log \frac{0.4}{0.6} \\ \implies & \frac{N_1}{N} > \frac{\log 5/6}{\log 2/3} = \frac{\log 1.2}{\log 1.5} = 0.4497 \text{ because } \log 2/3 \text{ is negative} \end{aligned}$$

Therefore, we have

$$\hat{\theta}_{MAP} = \begin{cases} 0.4 & \text{if } \frac{N_1}{N} < \frac{\log 1.2}{\log 1.5} \\ 0.5 & \text{if } \frac{N_1}{N} > \frac{\log 1.2}{\log 1.5} \end{cases}$$

Note that N_1/N can never be exactly equal to $\frac{\log 1.2}{\log 1.5}$ because $\frac{\log 1.2}{\log 1.5}$ is irrational.